Spencer Palladino
Stat 4220
Final Project
ALL CODE IS ATTACHED IN R FILE
Question 1:

(a) Estimate the factorial effects. Which effects appear to be important?

A	B	C	D	AB	AC	BC	ABC	AD	BD	ABD	CD	ACD	BCD	ABCD	Totals
-	-	-	-	+	+	+	-	+	+	-	+	-	-	+	13.413
+	-	-	-	-	-	+	+	-	+	+	+	+	-	-	29.926
-	+	-	-	-	+	-	+	+	-	+	+	-	+	-	23.724
+	+	-	-	+	-	-	-	-	-	-	+	+	+	+	35.088
-	-	+	-	+	-	-	+	+	+	-	-	+	+	-	20.554
+	-	+	-	-	+	-	-	-	+	+	-	-	+	+	8.466
-	+	+	-	-	-	+	-	+	-	+	-	+	-	+	18.613
+	+	+	-	+	+	+	+	-	-	-	-	-	-	-	26.363
-	-	-	+	+	+	+	-	-	-	+	-	+	+	-	17.512
+	-	-	+	-	-	+	+	+	-	-	-	-	+	+	33.919
-	+	-	+	-	+	-	+	-	+	-	-	+	-	+	27.534
+	+	-	+	+	-	-	-	+	+	+	-	-	-	-	39.463
-	-	+	+	+	-	-	+	-	-	+	+	-	-	+	24.183
+	-	+	+	-	+	-	-	+	-	-	+	+	-	-	12.029
-	+	+	+	-	-	+	-	-	+	-	+	-	+	-	22.125
+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	30.706

Col	A	B	C	D	AB	AC	BC	ABC	AD	BD	ABD	CD	ACD	BCD	ABCD	
Effect	3.02	3.98	-3.6	1.96	1.93	-4.01	0.096	3.14	0.08	0.05	0.10	-0.08	0.02	0.04	0.01	

FactorialEff(A) = ((29.926 + 35.088+ 8.466 + 26.363 + 33.919 + 39.463 + 12.029 + 30.706) - (13.413+23.724+20.554+18.613+17.512+27.534+24.183+22.125))/16 = 3.018875

FactorialEff(B)= ((23.724 + 35.088+ 18.613 +26.363+ 27.534 + 39.463 + 22.125 + 30.706) - (13.413+29.926+20.554+8.466+17.512+33.919+24.183+12.029))/16 = 3.975875

FactorialEff(ABCD)= ((13.413 + 35.088+ 8.466 +18.613 + 33.919 + 27.534 + 24.183 + 30.706) - (29.926+23.724+20.554+26.363+17.512+39.463+12.029+22.125))/16 = 0.014125

I did this same strategy for each factorial effect taking the positive totals and subtracting by the negative totals and dividing by 16. Based on my answers, I would say that A, B, C, D, AB, AC, and ABC are significant effects given how large the effects are.

(b) Use an analysis of variance or regression to confirm your conclusions for (a).

Based on our regression model, it looks like we can confirm our answers to part (a),

A, B, C, D, AB, AC, and ABC are significant effects

```
> eff = 2*mod$coefficients[-1]
> eff
           alloy$A                    alloy$B                        alloy$C
          3.018875                   3.975875                       -3.596250
           alloy$D              alloy$A:alloy$B                alloy$A:alloy$C
          1.957750                   1.934125                       -4.007750
   alloy$B:alloy$C              alloy$A:alloy$D                alloy$B:alloy$D
          0.096000                   0.076500                        0.047250
   alloy$C:alloy$D      alloy$A:alloy$B:alloy$C        alloy$A:alloy$B:alloy$D
         -0.076875                   3.137500                        0.098000
alloy$A:alloy$C:alloy$D  alloy$B:alloy$C:alloy$D alloy$A:alloy$B:alloy$C:alloy$D
          0.019125                   0.035625                        0.014125
```

```
Call:
lm.default(formula = alloy$Length ~ alloy$A * alloy$B * alloy$C *
    alloy$D)

Residuals:
    Min      1Q  Median      3Q     Max
-0.3305 -0.1500  0.0000  0.1500  0.3305

Coefficients:
                                   Estimate Std. Error t value Pr(>|t|)
(Intercept)                       11.988062   0.050361 238.042  < 2e-16 ***
alloy$A                            1.509438   0.050361  29.972 1.74e-15 ***
alloy$B                            1.987938   0.050361  39.474  < 2e-16 ***
alloy$C                           -1.798125   0.050361 -35.705  < 2e-16 ***
alloy$D                            0.978875   0.050361  19.437 1.49e-12 ***
alloy$A:alloy$B                    0.967062   0.050361  19.203 1.79e-12 ***
alloy$A:alloy$C                   -2.003875   0.050361 -39.790  < 2e-16 ***
alloy$B:alloy$C                    0.048000   0.050361   0.953   0.355
alloy$A:alloy$D                    0.038250   0.050361   0.760   0.459
alloy$B:alloy$D                    0.023625   0.050361   0.469   0.645
alloy$C:alloy$D                   -0.038438   0.050361  -0.763   0.456
alloy$A:alloy$B:alloy$C            1.568750   0.050361  31.150 9.49e-16 ***
alloy$A:alloy$B:alloy$D            0.049000   0.050361   0.973   0.345
alloy$A:alloy$C:alloy$D            0.009563   0.050361   0.190   0.852
alloy$B:alloy$C:alloy$D            0.017813   0.050361   0.354   0.728
alloy$A:alloy$B:alloy$C:alloy$D    0.007062   0.050361   0.140   0.890
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2849 on 16 degrees of freedom
Multiple R-squared:  0.9977,    Adjusted R-squared:  0.9956
F-statistic:   469 on 15 and 16 DF,  p-value: < 2.2e-16
```

(c) Write down a regression model that can be used to predict crack length as a function of the significant main effects and interactions you have identified.

Crack Length = 11.988062 + 1.509438 (A) + 1.987938 (B) - 1.798125 (C) + 0.978875 (D) + 0.967062(A*B) - 2.003875(A*C) + 1.568750 (A*B*C)
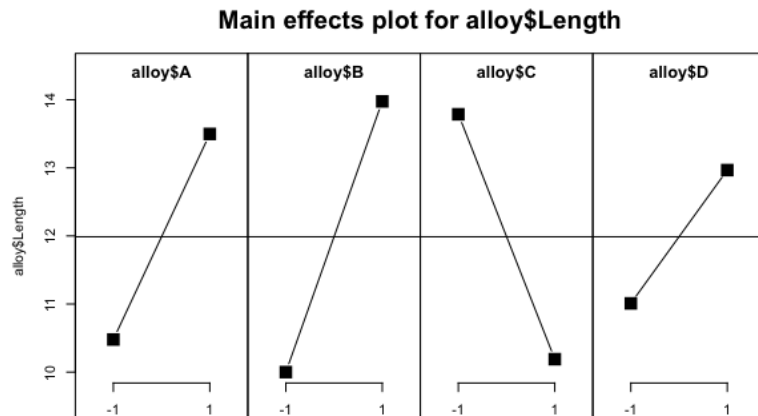
(d) Analyze the residuals from this experiment.

The residuals are randomly scattered and have a consistent variance for all fitted values so we can assume homosckedacity. The normal QQ plot brings some concern of how normal the distribution is.

Residuals vs Fitted

Normal Q-Q
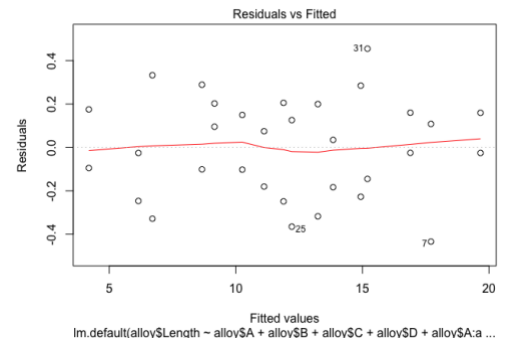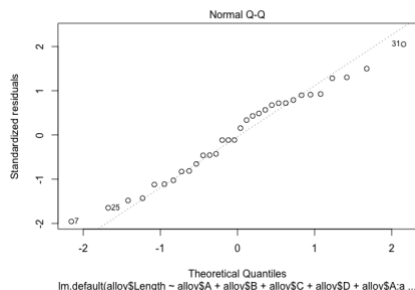
(e) What recommendations would you make regarding process operations? Use interaction and/or main effect plots to assist in drawing conclusions

Given that all main effects are significant, I would propose to use a lower pouring temperature, lower titanium content, and lower amount of grain refiner as well as a higher heat treatment method to get the smallest crack length possible based on this data. You can see this looking at the main effect plot slopes for each treatment as well as the model from part c.

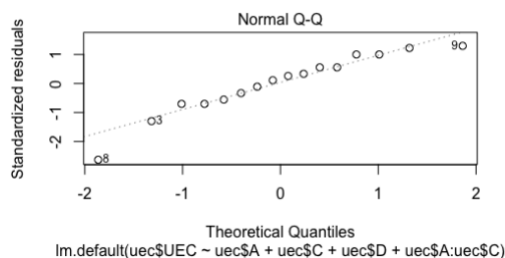**Main effects plot for alloy$Length**

Question 2:

(a) Analyze the data from this experiment. Which factors significantly affect UEC?

After analyzing the data and making my half normal plot, it would appear that factors A, C, D, and AC are significant.
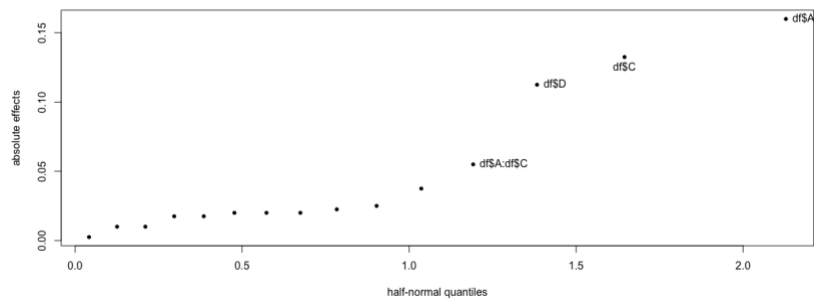

(b) Analyze the residuals from this experiment. Are there any indications of model inadequacy?

Looking at our residuals, I do not believe there is an indication of model inadequacy however I would look at sample #8 to double check just because it has a large residual and isn't on the QQ line.

Normal Q-Q
lm.default(uec$UEC ~ uec$A + uec$C + uec$D + uec$A:uec$C)

Residuals vs Fitted
lm.default(uec$UEC ~ uec$A + uec$C + uec$D + uec$A:uec$C)
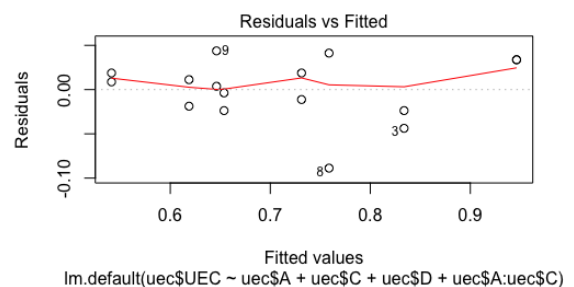
(c) Using the data construct and analyze a design in two blocks with ABCD confounded with blocks.

I made a 2 block design with B1 = AB and B2 = CD to confound with ABCD. I thought the residuals looked good and B2 to be potentially important.

Residuals vs Fitted

lm.default(uec$UEC ~ uec$A:uec$B + uec$C:uec$D)
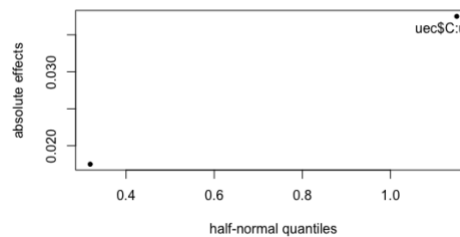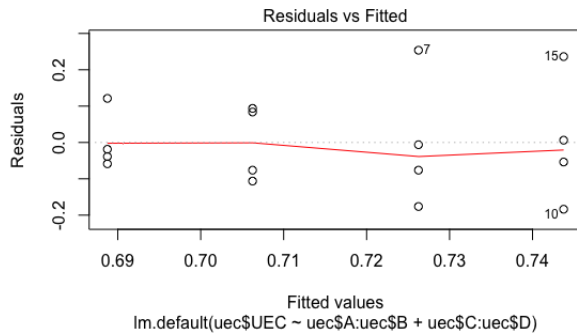
Residuals:
```
    Min      1Q  Median      3Q     Max
-0.18375 -0.07625 -0.02875 0.08625 0.25375
```

Coefficients:
```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.71625    0.03444  20.796 2.32e-11 ***
uec$A:uec$B  0.00875    0.03444   0.254    0.803
uec$C:uec$D  0.01875    0.03444   0.544    0.595
---
```

Question 3:

(a) Write out the alias structure for this design. What is the resolution of this design?

I=*ABCD* / Resolution IV

A = BCD / B = ACD / C = ABD / D = ABC / E = ABCDE / AB = CD / AC = BD / AD = BC / AE = BCDE / BE = ACDE / CE = ABDE / DE = ABCE

(b) Analyze the data. What factors influence the mean free height?

Looking at my initial summary of my data, it shows that factors A, B, E, and BE have significant impacts on the mean free heights. Factor D looks like it make be significant but at an alpha = .05 level it is not.

```
lm.default(formula = y ~ A + B + C + D + E + A:B + A:C + A:D +
    A:E + B:C + B:D + B:E + C:D + C:E + D:E)

Residuals:
     1        2        3        4        5        6        7        8        9
 0.01896  0.04062 -0.03854 -0.02104 -0.02104 -0.03854  0.04062  0.01896 -0.01896
    10       11       12       13       14       15       16
-0.04062  0.03854  0.02104  0.02104  0.03854 -0.04062 -0.01896

Coefficients: (3 not defined because of singularities)
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  7.625625   0.018115 420.947 2.96e-08 ***
A            0.121042   0.018115   6.682  0.00684 **
B           -0.081875   0.018115  -4.520  0.02025 *
C           -0.024792   0.018115  -1.369  0.26462
D            0.045625   0.018115   2.519  0.08628 .
E           -0.119375   0.018115  -6.590  0.00711 **
A:B         -0.014792   0.018115  -0.817  0.47401
A:C          0.000625   0.018115   0.035  0.97464
A:D         -0.011458   0.018115  -0.633  0.57197
A:E          0.031875   0.018115   1.760  0.17672
B:C               NA         NA      NA       NA
B:D               NA         NA      NA       NA
B:E          0.076458   0.018115   4.221  0.02431 *
C:D               NA         NA      NA       NA
C:E         -0.016458   0.018115  -0.909  0.43058
D:E          0.019792   0.018115   1.093  0.35448
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.07246 on 3 degrees of freedom
Multiple R-squared:  0.9791,    Adjusted R-squared:  0.8956
F-statistic: 11.73 on 12 and 3 DF,  p-value: 0.03323
```
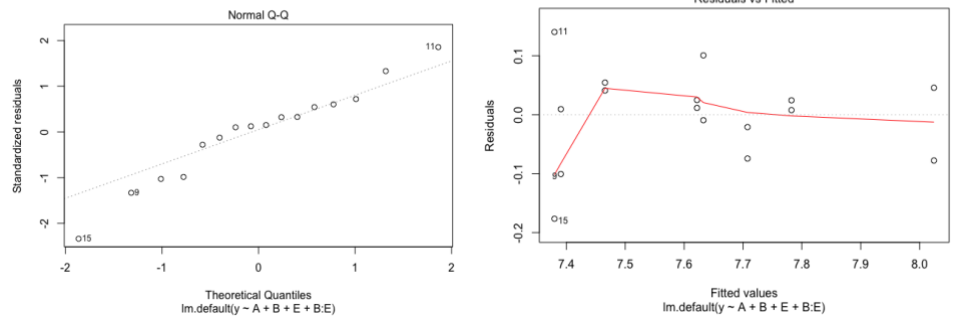
(c) Analyze the residuals from this experiment, and comment on your findings.

Normal Q-Q

Residuals vs Fitted

Looking at the residuals vs fitted graph, it looks alright but I am a little worried that the variance of the residuals is not equal for all x values but the normal qq plot looks reasonable.

(d) Is this the best possible design for five factors in 16 runs? Specifically, can you find a fractional design for five factors in 16 runs with a higher resolution than this one?

No this is not the best possible run for 5 factors in 16 runs because u could do a 2^5 factorial design with a resolution of V by setting the generator equal the highest order interaction which would be *ABCDE*

Question 4:

(a) Write the complete defining relation for this design. What is the resolution?

Resolution III

A=1, B=2, C=3, D=4, E=5, F=6, G=7, H=8, K=9

I = AFK = BGK = CHK = DEK = ADEF = BDEG = CDEH = ABFG = ACFH = BCGK = CEFB = BEFH = AEGH = DFHK = ABCE = BCDF = ACDG = ABDH = ABCDK = BCEFK = ACEGK = ABEHK= CDFGH = BDFHK = ADGHK = EFGHK= ABCDEFGH = ABDEFGK = ACDEFHK =BCDEGHK = ABCFGHK

(b) Write out the alias structure for this design (ignoring 3-factor and higher order interactions).

A = BJ = CH = DF = EG   /   B = AJ   /   C = AH,D = AF,E = AG   /   F = AD,G = AE /H = AC   /J = AB   / BC = DG = EF = HJ   / BD = CG = EH = FJ   / BE = CF = DH = GJ   /   BF = CE = DJ = GH   / BG = CD = EJ = FH   /   BH = CJ = DE = FG

(c) Analyze the data by fitting a main effects model. What factors appear to be important? Analyze the residuals and comment on your findings.

After fitting a main effects model and plotting the effects on a half normal plot there is not a clear distinction which effects are important because many of the p-values are not less than .05 but I want to say D, C, F, E, and maybe even A are important based on the half normal plot. To be more lenient and compromise the half normal plot, I will use an alpha = .15 and say D, C, F, and E are important factors.

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  75.3125     1.4662  51.366 3.65e-09 ***
A             2.1875     1.4662   1.492   0.1863
B            -0.8125     1.4662  -0.554   0.5995
C             3.4375     1.4662   2.344   0.0575 .
D             5.1875     1.4662   3.538   0.0122 *
E            -2.4375     1.4662  -1.662   0.1475
F             2.5625     1.4662   1.748   0.1311
G            -0.5625     1.4662  -0.384   0.7145
H            -1.9375     1.4662  -1.321   0.2345
K             0.5625     1.4662   0.384   0.7145
---
```
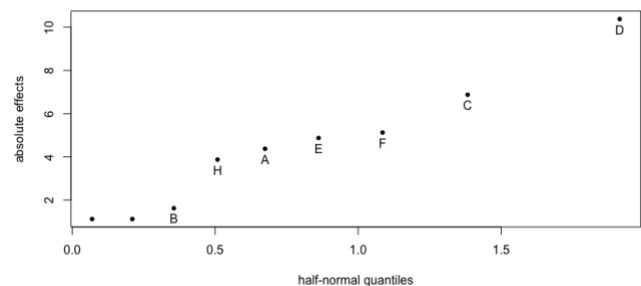

Using these factors, our residual plots look like so:
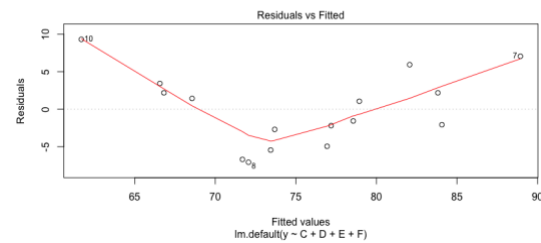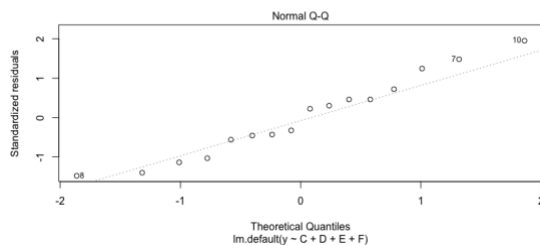
Based on these, I am not entirely happy with the residual vs fitted plot because of the 10th sample so I would want to look at that again. The QQ plot looks alright but once again the 10th sample stands out.

(d) Redo (c) by considering two-factor interactions besides main effects.

Involving so many confounding variables is quite difficult in r as seen by the summary to the right. This due to the issue of dfs. When using the 2 factor interactions of variables we found to be significant is realistic, however.

```
Coefficients: (21 not defined because of singularit
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  75.3125        NA      NA       NA
A:B           0.5625        NA      NA       NA
A:C          -1.9375        NA      NA       NA
A:D           2.5625        NA      NA       NA
A:E          -0.5625        NA      NA       NA
A:F           5.1875        NA      NA       NA
A:G          -2.4375        NA      NA       NA
A:H           3.4375        NA      NA       NA
A:K          -0.8125        NA      NA       NA
B:C          -1.1875        NA      NA       NA
B:D           0.5625        NA      NA       NA
B:E           0.1875        NA      NA       NA
B:F          -0.5625        NA      NA       NA
B:G           2.0625        NA      NA       NA
B:H          -2.5625        NA      NA       NA
B:K           2.1875        NA      NA       NA
C:D              NA         NA      NA       NA
C:E              NA         NA      NA       NA
C:F              NA         NA      NA       NA
C:G              NA         NA      NA       NA
C:H              NA         NA      NA       NA
C:K              NA         NA      NA       NA
D:E              NA         NA      NA       NA
D:F              NA         NA      NA       NA
D:G              NA         NA      NA       NA
D:H              NA         NA      NA       NA
D:K              NA         NA      NA       NA
E:F              NA         NA      NA       NA
E:G              NA         NA      NA       NA
E:H              NA         NA      NA       NA
E:K              NA         NA      NA       NA
F:G              NA         NA      NA       NA
F:H              NA         NA      NA       NA
```

When we do this we get a half normal plot that shows D:E, D:F, and C:D to be significant. Our residuals all look pretty good with a slight concern of increasing variability for higher x values.
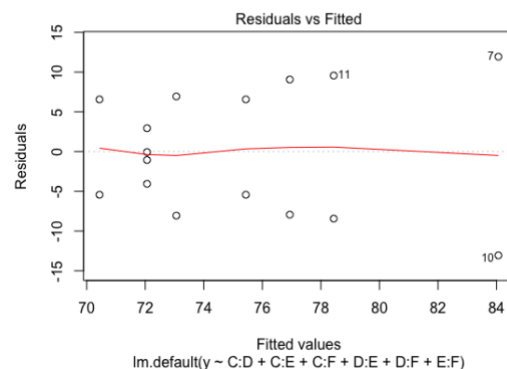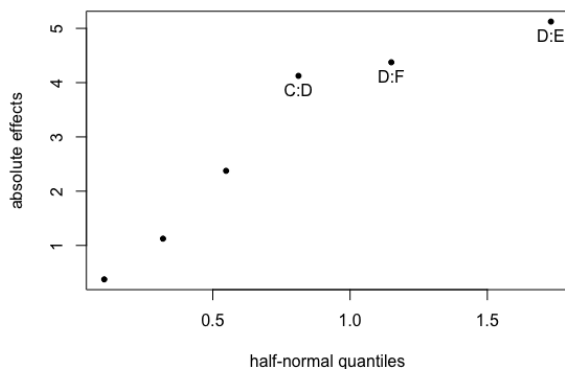
(e) The authors concluded that C, D, E, F, DE, DF were significant. Does your analysis support their conclusion? What assumptions did they make in order to conclude that DE and DF were significant? Do you agree?

I found C, D, E, F, D:E, D:F, and C:D to be significant so the only one we disagreed with was C:D. Generally, our conclusions are very similar and C:D was the worst of my 2 fi terms I accepted. In order to conclude DE and DF were significant they had to accept that affects that were aliased by them were not important variables. So for example A = DF, but I actually thought A might have been important but was not 100% sure. I would need to learn more about the compounds to actually disagree with them but based on my statistical knowledge I would question DF.